Bias in detrital fission track grain-age populations: Implications for reconstructing changing erosion rates

Bias in detrital fission track grain-age populations: Implications for reconstructing changing erosion rates

Mark Naylor[1], Hugh Sinclair[1], Matthias Bernet[2], Peter van der Beek[2], Linda Kirstein[1]

[1.] University of Edinburgh, School of GeoSciences, Edinburgh, UK

[2.] Universite Joseph Fourier, Institut des Sciences de la Terre, Grenoble, France

Corresponding author : mark.naylor@ed.ac.uk

Abstract

The sedimentary record is our principal archive of mass transfer across the Earth's surface in response to tectonic and climatic changes in the geologic past. The thermochronology of individual sediment grains (detrital thermochronology) has emerged as a critical tool to infer erosion rates and track mountain belt evolution. Such inferences are reliant upon the statistical inversion of detrital grain ages to unbiasdly approximate the cooling history of the source areas from which the sediment originated. However, it is challenging to critique the reliability and consistency of modelled ages. These arise both from fundamental measurement uncertainties and the assumptions we employ in inverting the data. For detrital fission track modelling of young detrital samples, this problem is particularly acute since the uncertainty on the track counts produces uncertainty in the age estimates. We apply Monte-Carlo modelling to generate synthetic detrital data conditioned on known closure age models, and then invert the grain data to assess the reliability of different inversion schemes. The results clearly demonstrate that existing practice can be subject to large uncertainty, to systematic bias and to non-uniqueness of interpretation. We then show how to map such regions of systematic bias in the population modelling as a function of the true closure ages, and how this bias propagates through into the lag-time modelling. Applying the method to real data from the Siwalik group sediments in western Nepal, we find no evidence for a

change in the underlying climate or tectonic processes, since the apparent change in lag coincides with a thresholded change in the resolution of the population modelling. This paper shows how to map regions of systematic bias in the population modelling as a function of the true closure ages, and how this bias propagates through into the lag-time modelling and can be applied retrospectively to existing studies. However, it is equally applicable to other age inversion schemes such as minimum age modelling. The application of these methods will enhance current practice and facilitate more robust interpretation of grain ages, in particular in discriminating between stationary and non-stationary geological and climatic processes.

1 Introduction

The reconstruction of past erosion rates is critical in determining the evolution of past sediment fluxes (Allen et al., 2014), the development of active mountain ranges (Jamieson and Beaumont, 1989) and evolving surface topography (England and Molnar, 1990). Erosion records the interplay of climate, lithology and tectonics, and so past erosion rates are also commonly interpreted in terms of these controls. Methodologies for reconstructing erosion over millions of years are dominated by measurements of sediment volumes and measuring the cooling history of rocks (thermochronology) where cooling is used as a proxy for exhumation of rock through erosion (Reiners and Brandon, 2006).

Bedrock thermochronology analyses the cooling history of a number of crystals from a single rock sample where all the grains have experienced the same history. In order to reconstruct the cooling history of a bedrock sample, multiple thermochronometers that record the time of passage through a range of closure isotherms are required. For regional analyses, multiple samples have to be collected, ideally from a range of different elevations (e.g. Fitzgerald et al., 1995); spatial interpolation may then enable regional erosion histories to be reconstructed (e.g. Vernon et al., 2008). Bedrock thermochronology often misses the early, now substantially eroded part of the exhumation history.

An alternative to regional bedrock sampling, is to analyse the sediment sourced from river catchments that drain the region over variable timescales; this approach is termed detrital thermochronology (Garver et al., 1999). In this method, grains originate from erosion of a source area that covers a broad region where exhumation rates are likely to vary, and hence the age distributions should record that variability. For sediment samples taken from the stratigraphic record, the age distributions record exhumation rates averaged over the time intervals for each grain to pass from its closure depth to the surface; these intervals will be different for the different populations. This ability to reconstruct ancient exhumation rates of the upper few kilometres of the crust for different stratigraphic time intervals using a single technique has increased understanding of the evolution of mountain chains such as the Alps (Bernet et al., 2009; Glotzbach et al., 2011) and the Himalaya (e.g. van der Beek et al., 2006).

A widely used thermochronometer is fission-track (FT) analysis of apatite and zircon grains (e.g., Gallagher et al., 1998; Tagami and O'Sullivan, 2005). In contrast to noble-gas based methods (e.g. Ar-Ar, (U-Th)/He), FT analysis is relatively insensitive to abrasion of grains during sediment transport and so is most suited to detrital thermochronology. In the simplest case, the FT method takes grains of apatite or zircon, measures the density of spontaneous fission tracks that have damaged the crystal lattice of the grain, and uses independent information on the amount of uranium in the grain to estimate a duration since the grain started accumulating tracks (Fleischer et al., 1975; Price and Walker, 1963). The detrital fission-track (DFT) method is similar to bedrock FT up to the point where individual grain ages are estimated. Since detrital samples contain grains from multiple bedrock sources that were exhuming at different rates, we cannot assume that a common pooled age is a useful measure for characterizing the composite cooling histories recorded in the detrital sample.

There are two commonly applied methods used to analyse detrital grain age data. The first uses population modelling to invert samples with up to ~120 grains sampled from modern river sediments or the stratigraphic record to find a parsimonious set of population ages that explain the variance in the observed grain ages (Brandon, 1992; 2002; Galbraith, 1988; Galbraith and Green, 1990; Galbraith and Laslett, 1993). These populations, especially the youngest P_1 population, are then interpreted to make inferences about the evolution of regional erosion rates, as well as sediment provenance and catchment reorganisation (e.g., Bernet et al., 2001; 2009; Glotzbach et al., 2011; Kirstein et al., 2010). The second method uses one of several methods to invert for a minimum age consistent with the population of ages (e.g. Galbraith and Laslett, 1993; Galbraith, 2005).

This paper considers how the component age populations in detrital fission-track analysis are estimated, and how they relate to "true" closure ages experienced by the grains. We explore issues of bias, uniqueness, and uncertainty when identifying component populations within detrital samples. Monte-Carlo sampling is used to generate synthetic DFT samples where the underlying closure age distribution is specified and our ability to recover the known closure age(s) in multiple random samples is tested. In these samples, the only sources of uncertainty are the probabilistic assignment of closure ages to each grain and Poisson counting errors on the number of spontaneous tracks. We then apply these methods to facilitate a more robust reinterpretation of detrital data from the Siwaliks (van der Beek et al., 2006). The power of this computational method comes in the scale of the samples that can be analysed; we have typically run ~500,000 synthetic detrital samples which takes ~2 days on a desktop computer. This allows us to critique previous interpretations of data in a transparent and systematic manner. Specifically, the method quantifies the emergence of artefacts as the resolution of population modelling degrades.

2 Fission Track Theory and Age Models

Fission tracks in crystals, such as apatite and zircon, are the record of lattice damage generated by the spontaneous decay of ^{238}U or the induced decay of ^{235}U. These tracks become visible under a microscope when the crystals are mounted, cut, polished and etched (Fleischer et al., 1975; Price and Walker, 1963). Annealing of the crystals removes fission tracks at a rate that increases with temperature; above some mineral-specific closure temperature all tracks are rapidly lost and below that temperature tracks accumulate. In un-reset samples, the spontaneous tracks provide a natural record of how many fission decays have occurred since cooling below the annealing temperature. In the external detector method (Gleadow, 1981), irradiation of the crystals induces ^{235}U to produce a set of induced tracks onto an external mica sheet from which the concentration of ^{235}U in the original crystal can be estimated. Since the isotopic ratio of ^{235}U/ ^{238}U is known, the induced track data can then be used to estimate the current concentration of ^{238}U in each crystal. Hence, independent knowledge of the amount of radioactive isotopes in a crystal and the measurement of the number of spontaneous tracks, which have only accumulated since cooling below the closure temperature, allows us to estimate how long the mineral has been below its closure temperature. These mineral-specific ages are known as grain ages and are calculated using the FT age equation (Price and Walker, 1963). When this equation is calibrated using the zeta method (Hurford and Green, 1983), the age estimate of the u^{th} grain is given by:

$$\tau_u = \frac{1}{\lambda_{U^{238}}} \ln\left(1 + \frac{1}{2}\lambda_{U^{238}} \zeta \rho_d \frac{\rho_{s,u}}{\rho_{i,u}}\right) \quad (1)$$

Where zeta, ζ is an empirically determined analyst-specific constant estimated by dating samples of known age (Hurford and Green, 1983). Other terms are the U^{238} decay rate ($\lambda_{U^{238}} = 1.55125 \times 10^{-4}\ \mathrm{Ma}^{-1}$), the track density in the standard glass used to measure the neutron fluence (ρ_d) and the ratio of spontaneous to induced track densities ($\rho_{s,u}/\rho_{i,u}$).

When measured over a fixed area, the density ratio can be estimated from the ratio of the number of tracks $N_{s,u}/N_{i,u}$.

Due to the effect of counting errors on $N_{s,u}$ and $N_{i,u}$, individual grain ages are often subject to large uncertainties. Hence, the motivation for studying samples composed of many grains is to enhance the statistical robustness by finding a set of age populations consistent with a statistically well sampled set of observed grain ages. This problem is more acute for detrital apatite fission-track analysis than for detrital zircon. The latter system is characterized by a higher closure temperature and generally higher U-content; thus both $N_{s,u}$ and $N_{i,u}$ will tend to be higher for zircon (e.g., Garver et al., 1999).

2.1 Inverse Models for Fission Track Ages

Inverse modelling refers to the numerical process by which information is obtained from a (detrital) FT sample with the intention of enabling an interpretation to be made regarding the cooling histories of the grains.

2.1.1 The pooled age

Where the Chi-squared test indicates low age dispersion, it can be assumed that all of the grains have experienced the same temperature history and hence the same true cooling age; this is commonly the case in bedrock samples. The maximum likelihood solution for the common pooled age is given by:

$$\tau_{pooled} = \frac{1}{\lambda_{U^{238}}} \ln\left(1 + \frac{1}{2}\lambda_{U^{238}}\zeta\rho_d \frac{N_{s,total}}{N_{i,total}}\right) \quad (2)$$

Where $N_{s,total} = \sum_{u=1}^{n} N_{s,u}$ and $N_{i,total} = \sum_{u=1}^{n} N_{i,u}$ are the sum of all the spontaneous and induced tracks measured across all the grains.

2.1.2 The central age and age dispersion

In detrital samples, it is rare that the grains will represent a common age unless they have been heated during burial; low χ^2 values (<5%) and large grain age dispersion are used to

justify more sophisticated analysis. The central age (Galbraith and Laslett, 1993) generalizes the pooled age by relaxing the assumption that all of the grains come from a single true age with no variance (i.e. $\rho_s/\rho_i = \text{constant}$). The central age uses a random effects model where $\log(\rho_s/\rho_i)$ is drawn from a normal distribution. The population mean of $\log(\rho_s/\rho_i)$ is referred to as the central age and the standard deviation of the distribution, σ is the age dispersion (e.g. Galbraith, 2005). The central age is of limited use when the detrital sample is composed of distinctly different components.

2.1.3 Component age populations

Equation (1) can be used in conjunction with an appropriate inversion scheme to find a parsimonious set of closure age populations, P_j that explain the observed variance in the individual grain ages (Brandon, 1992; 2002; Galbraith, 1988; Galbraith and Green, 1990; Galbraith and Laslett, 1993). These principal age components, especially the youngest P_1 component, are frequently used to interpret the exhumation of the source area using the 'lag-time' concept, ultimately used to interpret tectonic and climatic controls on past erosion rates.

The difference between a population age and its stratigraphic age is termed the 'lag-time'; when the estimated depth to the closure isotherm is divided by the lag-time, it yields an exhumation rate at the time of sediment deposition (Bernet et al., 2001; Garver et al., 1999; Garver and Brandon, 1994). Hence, the evolution of lag-times through a sedimentary succession sourced from a mountain range, yields information on changing exhumation rates (e.g. Bernet and Garver, 2005). However, the ability to identify unbiased representative age population(s) from the sample is a necessary pre-requisite for defining comparable lag-times.

Further, fitting an integer number of populations to what is often a relatively continuous distribution of individual grain ages introduces undesirable thresholds into the analysis (e.g., Vermeesch, 2007). There will be systematic shifts in the population ages when the number of

populations change, as resolution thresholds are crossed. This paper will examine the consequences of this on the interpretation of DFT data.

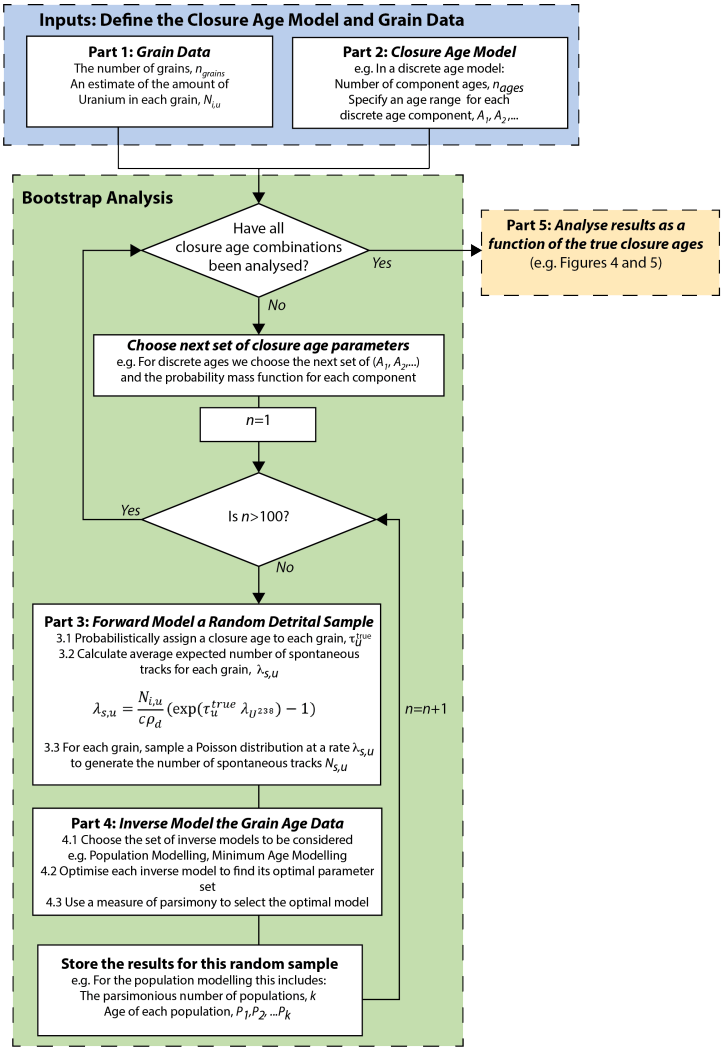


Figure 1 Flow chart of the bootstrap method for quantifying bias and uncertainty in the inverse age modelling of detrital fission track samples. The motivation is to generate synthetic detrital samples using known closure age models and explore how well different inversion schemes recover key attributes of the original closure age models. In addition to varying the closure age model, we also explore the impact of detrital samples consisting of different numbers of grains and with different uranium content.

3 Quantifying Uncertainty and Bias using the Monte-Carlo Bootstrap

The basis of the bootstrap method (Figure 1) is to (i) forward model a synthetic detrital fission track sample using a known closure age model and (ii) invert the resulting dataset to quantify

how well we can recover information about the original closure age model. By repeating this many times for the same age distribution, we identify systematic bias and determine uncertainty as a result of: 1) stochastic uncertainty on the number of observed tracks and, 2) fundamental limitations in the inversion scheme. By systematically sweeping through a range of closure ages, we map how uncertainty and bias vary through a detrital record.

3.1.1 Part 1: Grain Data (Fig. 1)

A detrital sample is composed of a number of grains, each of which has been analysed to determine an induced track density as a proxy for its uranium content. The grain specific uranium concentration limits the rate at which fission tracks can spontaneously decay and hence affects the quality of that grain as a chronometer. Empirically, the uranium content of grains in a detrital sample is over-dispersed (e.g. Figure 2), it would be erroneous to assume an average concentration. By using a measured set of induced tracks, $N_{i,u}$ as the input to the analysis, we can quantify sample specific bias and uncertainty.

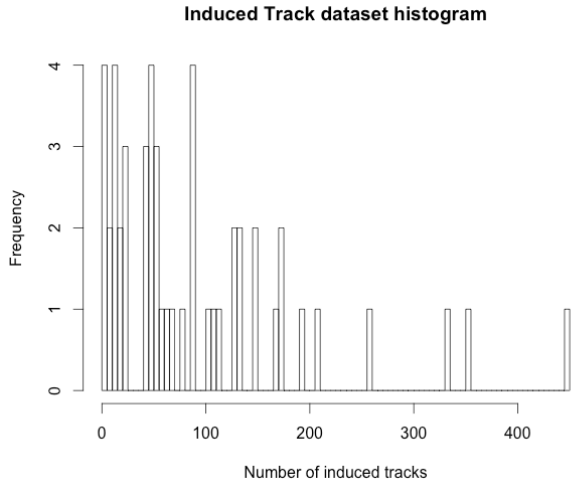


Figure 2 Histogram showing the number of induced tracks in an Alpine detrital sample using a 5-track bin-width. This set of 51 values is used in our initial analysis. These values are given in the supplemental material.

3.1.2 Part 2: Closure Age Model

The closure age model is used to probabilistically assign a closure age to each grain in a synthetic detrital sample. It is a simplified model for hypothetical true closure ages based on the geological history. This is in contrast to the inverted ages derived from a statistical analysis of a set of grain ages, which might not be able to resolve the true closure ages in practice. The functional form of the closure age model (i.e. it's general shape) is a subjective choice, and is informed by the geological scenarios we wish to compare.

For example, in this paper we focus on discrete closure age models that consist of a set of n discrete closure ages $A_1, A_2, \ldots A_n$, each with an associated probability that describes the proportion of grains in the bulk sample of that age $\pi_1, \pi_2, \ldots \pi_n$. Since these probabilities must sum to 1, there are $n-1$ independent probabilities. For the examples in this paper equipartition is assumed such that $\pi_1 = \pi_2 = \cdots = \pi_n$. So for 2 closure ages, there is a 50:50 split and for 3 closure ages the probability for any single component is 1/3. We will explore when population modelling cannot resolve the correct number of age components and quantify the impact of this on P1 age bias.

As an alternative example, we could choose a continuous distribution if we wanted to understand how informative population modelling is when detrital samples are derived from the convolution of hypsometry and an age-elevation profile.

As well as choosing the structure of the closure age model, we must also define ranges for the parameters in the closure age model. A pragmatic choice ensures that the model can explore over the range between the youngest and oldest grain ages in a fairly uniform manner.

3.1.3 Part 3: Forward Modelling a Synthetic Detrital Sample

The forward model probabilistically assigns a closure age to each grain τ_u^{true}, and rearranges the fission track age equation (Eqn 1) to calculates the average expected number of spontaneous tracks for each grain $\lambda_{s,u}$,

$$\lambda_{s,u} = \frac{N_{i,u}}{c\rho_d}(\exp(\tau_u^{true}\,\lambda_{U^{238}}) - 1)$$

Note that the mean expected number need not be an integer since it is the value measured on average if many replicates were available. However, measurements of $N_{s,u}$ are integer counts, thus the observed number of spontaneous tracks $N_{s,u}$ on each grain is a Poisson distributed random variable with the mean expected number of tracks being $\lambda_{s,u}$,

$$N_{s,u} = Pois(\lambda_{s,u})$$

Where there are short lag-times and/or where there are low track counts, the Poisson errors may lead to a significant number of grain ages that are younger than the depositional age. Thus, each synthetic detrital sample consists of a set of $N_{s,u}$ and $N_{i,u}$ for $u = 1, n$ grains, and using a specified ζ calibration; this is equivalent to the data that would be measured in a real detrital study. In addition, we know the true set of closure ages used to generate the synthetic data, $A_1, A_2, \ldots A_n$, the probabilities with which they are assigned $\pi_1, \pi_2, \ldots \pi_n$ and how they are assigned to the grains.

3.1.4 Part 4: Inverse Model the Synthetic Grain Data

A perfect inverse model would be able to take the detrital data and uniquely determine everything about the closure age history experienced by the grains in the sample; in practice this is not possible. The nature of the bias and uncertainty introduced during the inverse modelling is dependent upon the data quality, the nature of the features we want to resolve, the complexity of the true closure age distributions, and the statistical methods chosen to extract the relevant metrics.

In addition to choosing a statistical model (e.g. the population modelling or minimum age modelling), we also require a method to optimise the parameters for that model (e.g. Expectation Maximisation or Markov Chain Monte Carlo) and a statistical measure of parsimony to discriminate between competing models (e.g. ΔBIC, F-test).

We use the expectation maximisation method (See Supplemental Material) which estimates the maximum log likelihood, $\mathcal{L}_{max}$, for each set of synthetic data. An ensemble approach is used to reduce the risk of finding a local minima; this automation was necessary as we have typically analysed ~500,000 synthetic samples for each of the result summaries. Our conclusions are independent of the choice of algorithm applied to maximize likelihood, provided we are close to the global minimum; rather the functional form of likelihood function itself is critiqued. Put simply, the limits of what the likelihood function is capable of resolving when inverting for component closure ages in detrital FT samples is explored.

Once we have maximized the likelihood to find a best fitting set of parameters for competing models, we need a way to choose which of the models we prefer, e.g. how many peaks are fitted. Often, a statistical measure of parsimony is applied using either an F-test or applying a Bayesian method such as the Bayesian Information Criterion (BIC; Schwarz, 1978). These criteria overcome the problem of over-fitting the data by penalizing the quality of the fit as a function of the number of parameters used in each model and the number of observations being considered.

In the case of fitting multiple populations, the number of free parameters changes with the number of populations. For a single discrete age, only one parameter is needed, i.e. the P_1 age. For more populations additional information is required for each peak, in particular an occupation probability π_j, which is the proportion of grains that are included in the corresponding peak P_j. Since these probabilities must sum to 1 the number of independent

parameters for a k-age population model is $p = 1 + 2(k - 1)$. Knowledge of the geological context has been used to constrain k, or at least provide a minimum value (e.g. Kirstein et al., 2010). In practice, it is challenging to constrain the number of populations in advance of the analysis without biasing the interpretation.

3.1.5 Part 5: Analysis

Once 100 replicates have been generated for each set of closure age parameters, the inverted age distributions can be compared with the true closure age model parameters to assess systematic bias and quantify uncertainty. For example, as part of the assessment for population modelling of peaked closure age models, it should be considered whether the number of populations in the inverted solution is the same as the number specified when creating the sample. When using population modelling, it should be asked whether there is a systematic difference between the P_1 age and the known youngest closure age. Similarly, when doing minimum age modelling, do systematic differences exist between the inverted minimum age and the youngest age specified in the closure age model?

4 Results using Discrete Closure Age Models

We determine what closure age models are resolvable in principle using population modelling for a specific detrital sample. By specific, we mean a finite set of n grains with some measured uranium content, likely determined through the external detector method.

When discussing the results, the term “closure age” is reserved to refer to the ages that are probabilistically assigned when generating the synthetic track data; in this sense, they are the “true” ages on which the synthetic data are conditioned. Age “populations” are used to refer to the inverted ages in the parsimonious model. When the inversion algorithm works well, the latter will be an unbiased estimate of the former. However, this is not always the case, and we quantify this bias using synthetic closure age models where everything is known.

Firstly, we need assign the grain data. Since N_i is over dispersed, we pragmatically choose to use the measured set of 51 N_i presented in Figure 2 since this is representative of the range of N_i we might find in a real sample. For the examples where the number of N_i is increased this original set is duplicated; thus our samples contain either 51 or 102 grains.

As described above (Fig. 1), each grain, u is randomly assigned a true closure age (τ_u^{true}) from a set of discrete closure ages (A_i) to create a synthetic DFT sample. We then invert the sample using the population model and apply the BIC to identify the statistically parsimonious number of age components. These inverted age populations, P_j are then compared with the closure ages, A_i, to look for systematic biases.

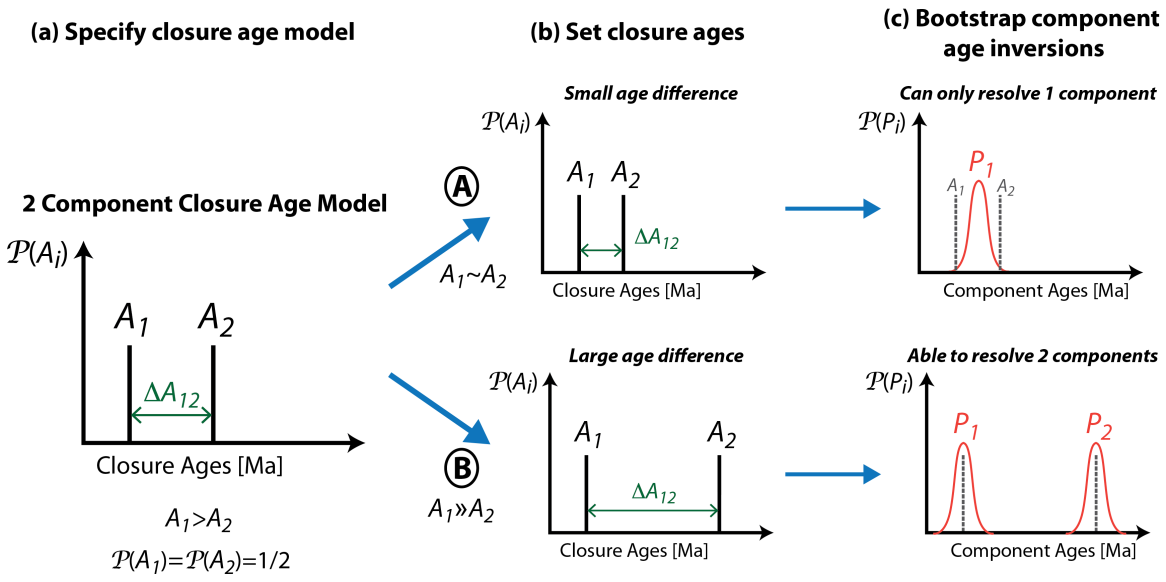


Figure 3 Cartoon illustrating the relation between the true closure ages (A_i) and inverted grain-age populations (P_j) for a synthetic sample conditioned on 2 discrete closure ages. In scenario (A) the age difference ΔA_{12} is too small for the inversion scheme to resolve the 2 components and the single P_1 age model is preferred. In scenario (B) the age difference ΔA_{12} is large enough for the inversion scheme to resolve 2 component ages with corresponding age estimates P_1 and P_2.

4.1 Bimodal Synthetic with 2 true closure ages

The simplest non-trivial closure age model has 2 discrete closure ages A_1 and A_2, where the difference between the ages is ΔA_{12}. In principle, the population modelling should perform well on this type of sample since the closure age model is functionally similar to the population model being fitting, i.e. both consist of a low number of discrete ages. However, the inversion scheme will find it harder to resolve closure ages when ΔA_{12} is small (Fig. 3Figure 3). We explore how the ability to resolve 2 populations varies as a function of A_1 and A_2.

The preference for the statistically parsimonious solution to contain 1 or 2 component populations as a function of A_1 and A_2 is shown in Figure 4. The *x*-axis shows the youngest closure age, A_1, which is varied from 1 Ma to 120 Ma in 1-Ma increments. Figure 4(a,b) show the second closure age, A_2 on the vertical axis. The diagonal trend is removed from these figures by plotting the relative age of the second component (i.e. $\Delta A_{12} = A_2 - A_1$) rather than the absolute value, which makes the plots easier to visualise. This relative age is plotted on the *y*-axis of Figure 4(c,d) and is varied across the range ΔA_{12}= 1 Ma to 40 Ma in 1-Ma increments. Each cell on the plot corresponds to the number of component populations averaged over 100 synthetic samples; green indicates preference for 2 components, red for one component with the transition region where counting errors in the track counts lead to variability in the preferred model. Figure 4 (a,c) used 51 grains and (b,d) used 102 grains in each sample. For these examples, the inversion scheme never preferred an age model with more than 2 populations.

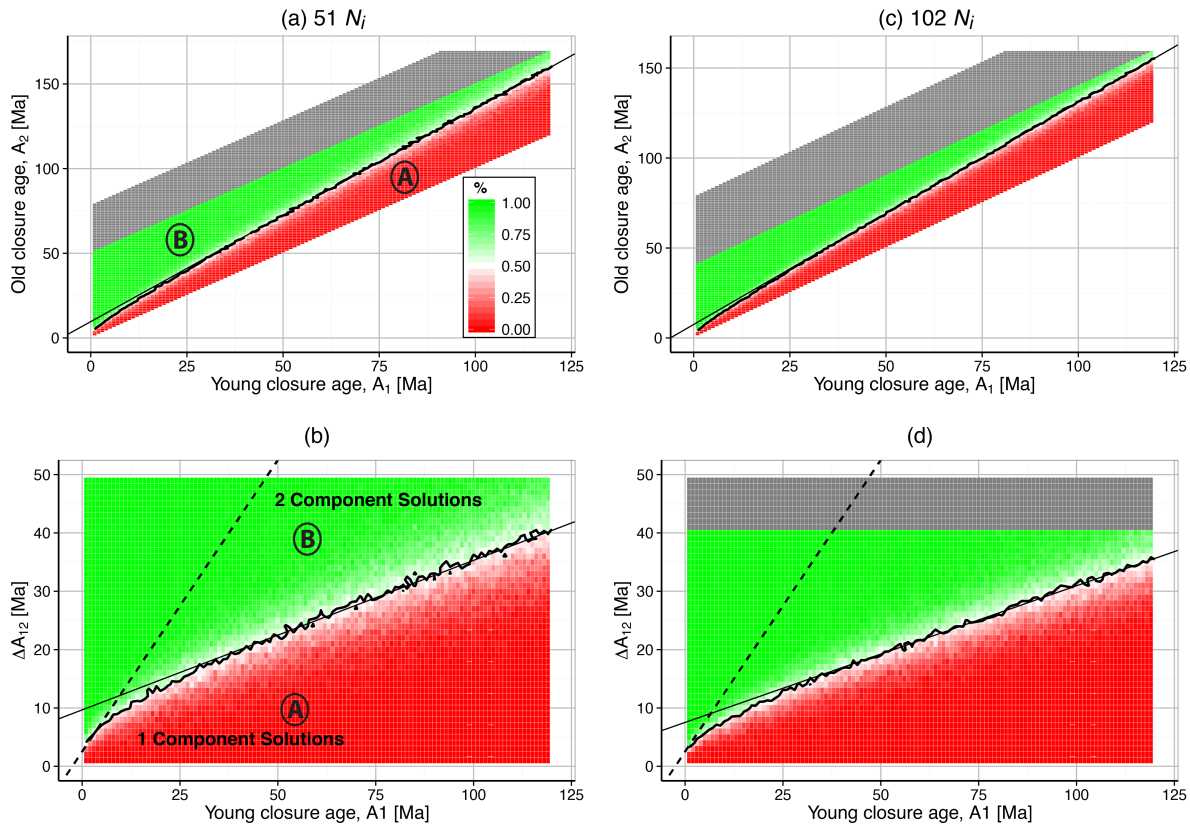


Figure 4 Image maps showing the preference for 1 (region (A) in red) or 2 (region (B) in green) component populations as a function of the true closure ages, A_1 and A_2. (a) and (b) plot the true closure ages on the *x* and *y*-axes and each point represents the proportion of 2 peak solutions averaged over 100 synthetic samples. (c) and (d) show the A_1 closure age on the lower axis and the relative difference to A_2 from A_1 age on the y axis; this format makes the plots significantly clearer and will be used in the rest of the paper. The wiggly black line shows the 50% contour, i.e. where there is an equal preference for 1 and 2 component solutions. The solid black straight line is a regression of the linear portion of this boundary. Increasing the number of grains increases the ability of the inversion to resolve the correct number of populations, but the difference is relatively small compared to the absolute ages. The black dashed line shows a 1:1 age line for reference. The ages derived from the models in (d) are analysed in Figure 5.

As an example of how to interpret this plot, consider the 102-grain case in the right hand column. At $A_1 = 1$ Ma we can resolve closure ages that are separated by a few million years (Figure 4d). However, for $A_1 = 55$ Ma the age difference required to resolve the closure ages needs to be closer to 20 Ma. The older samples require larger absolute age differences in order for them to be resolved.

The first order observations are:

(i) The ability to distinguish separate closure age components decreases as the absolute grain ages increase.

(ii) The line separating the one and two component population solutions is roughly linear for $A_1 > 25$ Ma for the examples shown.

(iii) For $A_1 < 25$ Ma, we can resolve populations at shorter lags than this linear trend would predict.

(ii) Increasing the number of grains in a sample allows us to resolve smaller closure age differences, although this tends to a finite limit.

When real datasets are analysed it is commonly assumed that the P_1 population is an unbiased estimate of the youngest true closure age A_1; in Figure 5 we examine whether this is a good assumption for the 102 grain case (Figure 4d). As for Figure 4, each cell is a summary statistic taken from the 100 model runs using that pair of closure ages. The left-hand column shows a summary of the all the P_1 population ages irrespective of whether a 1- or 2-component model is preferred; this is done because in real data we would not know how many true components there are, we would just take the P_1 ages and continue the analysis. The right hand column summarises the P_2 ages; these only exist when the 2-component models are preferred, which explains the lack of data in the lower right region of the graph where only a single population is chosen.

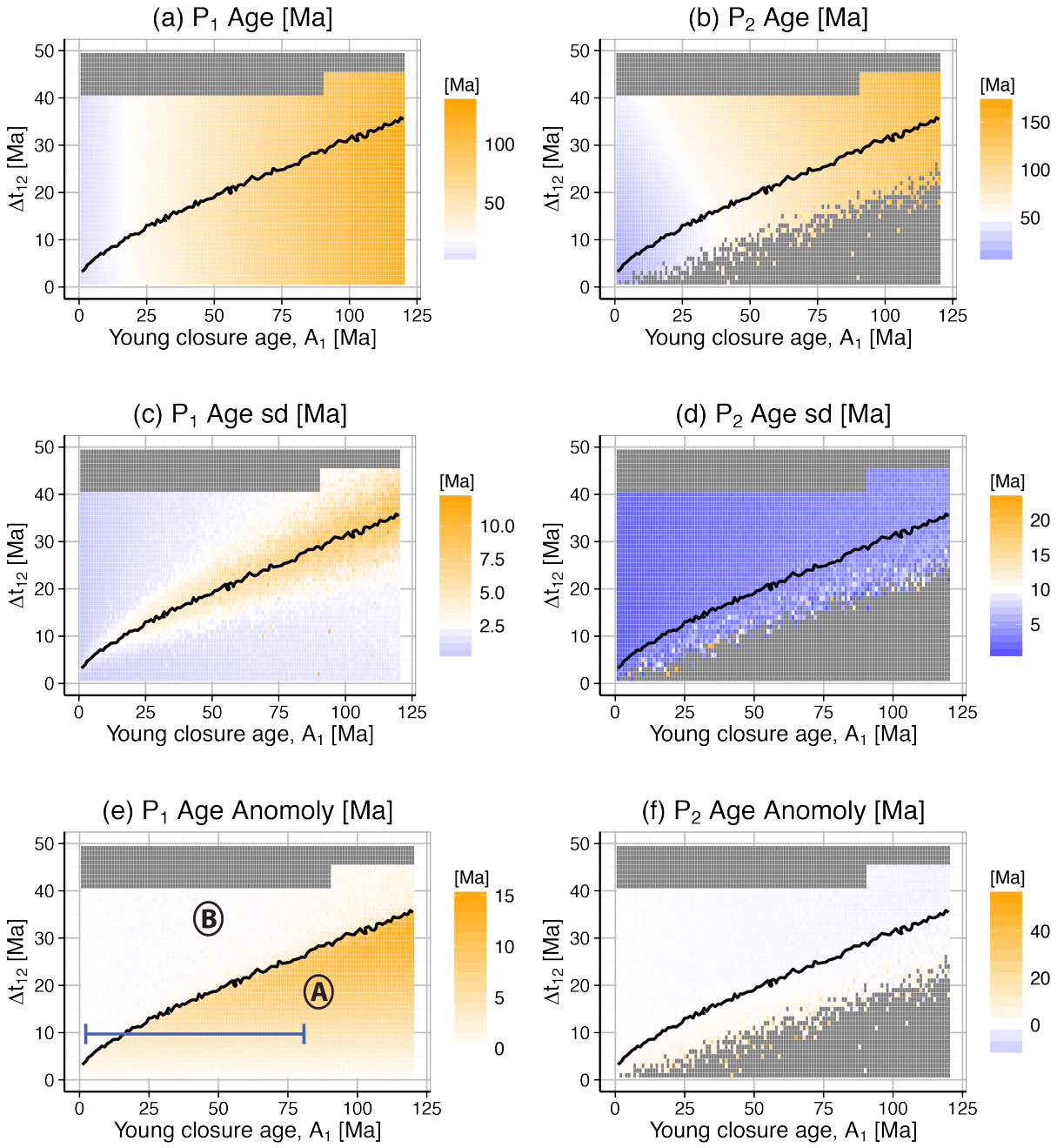


Figure 5 Analysis of the age populations associated with the preferred solutions of the bimodal closure age synthetics. This figure builds on the bimodal model presented in Figure 4d by presenting the corresponding P_1 and P_2 population age data. (a) and (b) show the median of the 100 inverted P_1 and P_2 ages for each age combination. The black line shows the boundary between preference for the 1 and 2 component population models derived from Figure 4. The standard deviation for the 100 ages is presented in (c) and (d) for the P_1 and P_2 ages. The uncertainty in P_1 ages increases about the boundary between the preference for 1 and 2 solutions. Due to the large range in median ages, it is easier to identify bias by plotting the population age anomaly, in (e) and (f), which is the difference between the median population age and the corresponding closure age used to generate the synthetic dataset. The P_1 age

anomaly shows a system step for unbiased ages in domain (B) where 2 populations are preferred to systematically older ages in domain (A) where 1 population solutions are preferred.

Figure 5(a,b) shows the median age of each population, $\overline{P_j(A_1, A_2)}$ in the 100 synthetic samples; as one would expect, to a first order, the inverted populations ages (P_1, P_2) scale with A_1 and A_2. We choose the median age as it is not strongly influenced by outliers. Figure 5(c,d) shows the standard deviation of the population ages; the P_1 component has a high standard deviation near the transition from 1 to 2 populations. The high standard deviation reflects the mixing of 1- and 2-age component populations in the inverted ages. In contrast, the spread in P_2 is relatively constant throughout since there is, by definition, only a P_2 component for the 2-component population model.

Figure 5(e,f) shows the population age anomaly (i.e. the median population age minus the true age on which the synthetic samples are conditioned, $\overline{P_i} - A_i$), which should be approximately zero where the age populations are unbiased estimates of the closure ages (pale colours). The anomaly plots easily identify bias in population age modelling; they show similar information Figure 5(a,b), but the colour scale is more sensitive when showing relative age differences. Orange indicates that the component population is older than the closure age on which it was conditioned; Blue indicates that the component population is younger than corresponding closure age.

In Figure 5(f), when $k = 2$, the P_2 age is an unbiased measure of A_2. In part (e), in the region where $k = 2$, P_1 is an unbiased measure of A_1; however, in the region where the method resolves a single population ($k = 1$ models) P_1 is systematically older than A_1. In fact, when $k = 1$, P_1 is an estimate of the central age of the bulk sample.

In summary, when the inversion scheme picks the wrong number of components it introduces systematic errors in the P_1 estimate of A_1. We now consider how propagating this leads to biased interpretation.

4.2 Lag-time plots and population modelling

Consider the idealised scenario of a single catchment that contains a significant structure, across which there is a change in the long-term erosion rate (e.g. Figure 6a) such that the difference in the lag times has been sustained at 10 Myr. For simplicity, assume that the sediment fluxes from these 2 regions mix in a 50:50 ratio. Thus the true closure ages of these 2 regions are plotted in Figure 6b. However, we have shown that the resolution of the population modelling reduces with the sample age, hence, in practice we will not be able to recover both true ages for older samples (Figure 6c).

We can start to see the impact of such a scenario, for a difference in closure ages of $\Delta A_{12} = 10$ Ma, by following the blue line on Figure 5(e). For young A_1, P_1 is a good estimate of the closure age of the faster eroding catchment and P_2 is a good estimate of A_2.

If the depositional age of the sample can be dated reasonably accurately, and independently of the DFT method, a lag-time plot can be constructed to investigate changes in lag-time. Each point in Figure 5(e) summarised the results of 100 synthetic samples, for the lag-time plot (Fig. 7Figure 7) we randomly draw a single P_1 age at each point along the blue line in 1 Myr intervals. Further, we assume it takes 1 Myr to go from passing through the closure temperature to being deposited in its modern stratigraphic position, though this assumption can be relaxed in practice.

The lag-time plot (Figure 7) shows a change in lag-time from ~ 3 Myr at young stratigraphic ages $(\text{for } A_1 < 9 \text{ Ma})$ to ~ 8 Myr at older stratigraphic ages $(A_1 > 21 \text{ Ma})$ despite there being no change in lag-time in the true closure age models. This change in the lag-time could

falsely lead to an interpretation of a reduction in erosion rates during the period $\sim ca.\,22 - 10$ Ma. The sharpness of this transition will be dependent upon the sampling density across the transition region; if there are only a few measurements across this window, as is more typical of real samples, the transition may appear sharp when in fact it is not.

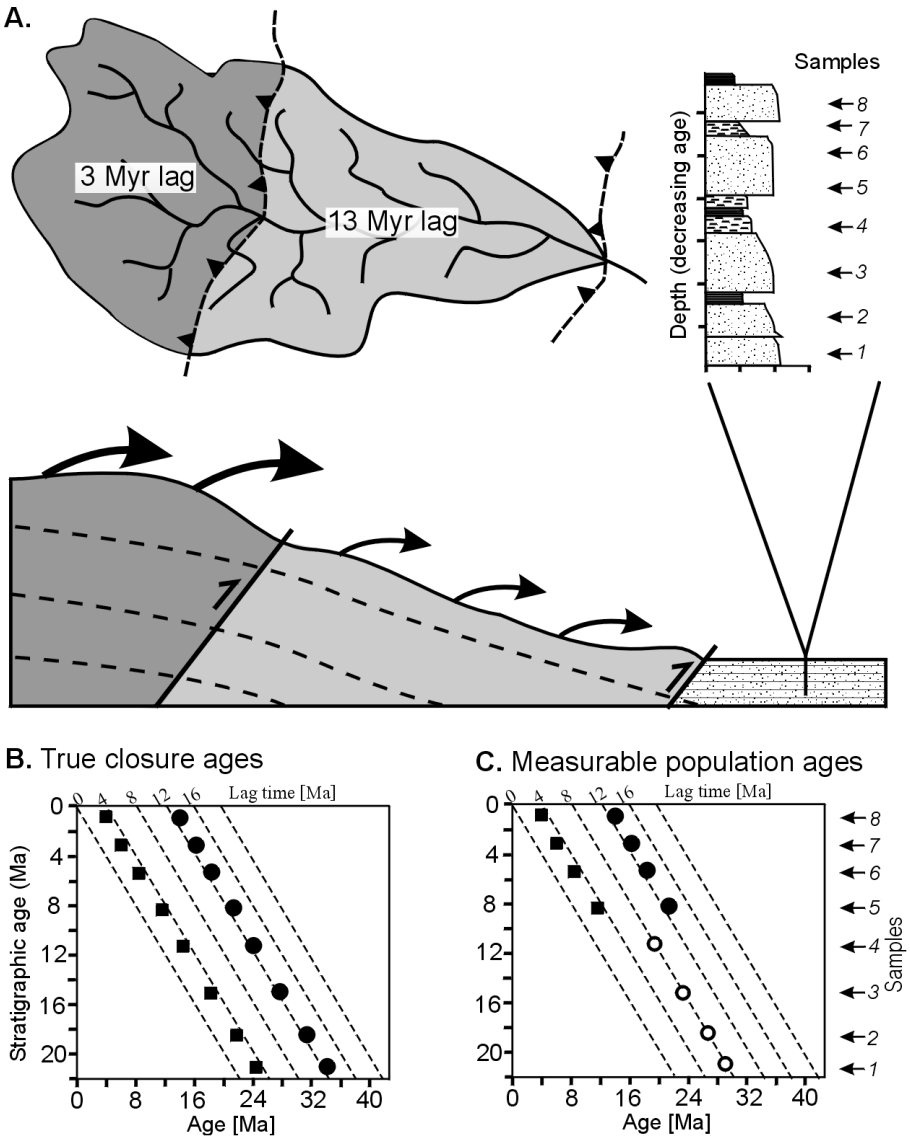


Figure 6 Cartoon illustration of how we might construct an analysis motivated by a simple geological model. (a) A catchment contains a significant structure, across which there is a change in the long-term erosion rate. The erosion rates translate into 2 different lag time of a 3Myr and 13Myr to deposition. We consider what the modelled thermochronometric ages might look like as a function of the stratigraphic age of the sample. (b) The 'true' closure ages are the lags experienced by the detrital samples – hence there are 2 independent sources represented at each stratigraphic age. Solid squares show the 3Myr lag and solid circles the 13Myr lag. (c) However, we have shown that the resolution of the detrital fission track methods reduces for older samples, so at some point we lose the ability to resolve both populations

when using a population modelling approach. This loss of resolution is indicated by the transition from filled symbols to a central age shown with the hollow circle. Where this transition occurs in practice will be a function of the number of grains, the amount of uranium, the difference in the true ages as well as a random component derived from counting errors.

The origin of this artefact is clear; a systematic change in the number of populations that can be resolved by the population modelling results in a systematic change in the age of the P_1 population (Figure 5e and 6c).

A simple check on whether changes in the number of populations can explain apparent trends in the lag-time plots, is to plot the number of populations as the symbol in the lag-time plot (e.g. Figure 7) since this information is already available during the standard analysis. This shows that the change in P_1 age is associated with a change in the number of component populations and could therefore be an artefact associated with changing resolution. This plot also highlights how broad the transition region is and prompts questions as to whether fluctuations over this $\sim 12\ \mathrm{Myr}$ period are explainable in terms of geology rather than statistical fluctuations derived from the inversion process. Whether there is sufficient data to resolve the breadth of the transition region depends upon the stratigraphic sampling density; low sampling may make a transition appear sharper than it really is.

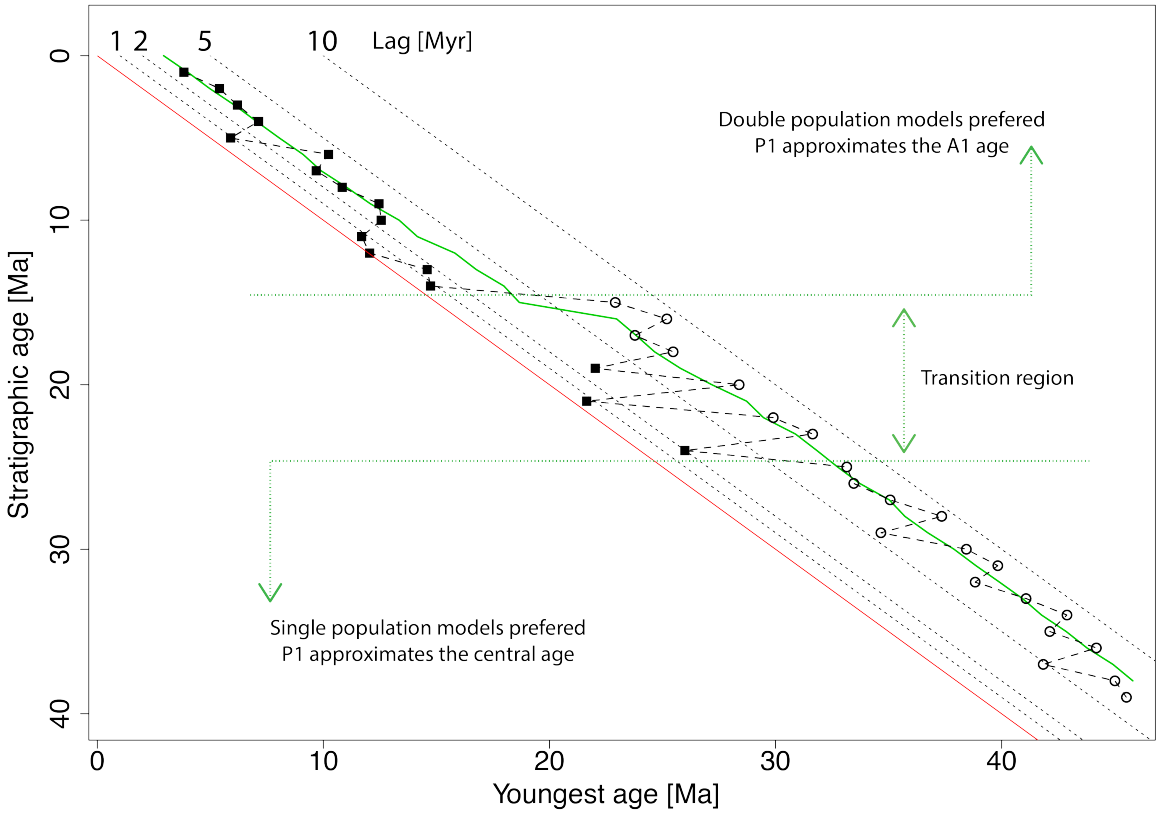


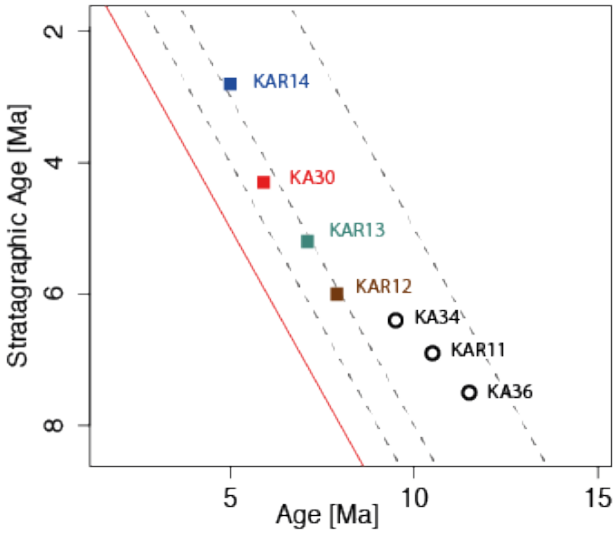
Figure 7 Lag-time plot for the bimodal closure-age model where the synthetic age model is conditioned with a constant age difference of $\Delta A_{12} = 10$ Ma. The symbols are randomly chosen youngest ages from each set of 100 synthetic models; the black squares indicate that the chosen age was a P_1 age within a 2 population model; the hollow circles indicate that the preferred solution was a single central age. The green line shows the median of the 100 youngest ages. The red line shows the depositional age, the true closure ages must be older than this. The dashed black lines show age lags at 1, 2, 5 and 10 Myr behind the depositional age. At young depositional ages, the P_1 is a reasonable estimate of the true closure age which is at a lag of 3Myr. At ages older than ~22Ma, we systematically get ages ~8Myr older than the depositional age. There is a sharp transition in the median age from the A_1 to the pooled age at stratigraphic ages of ~15Ma. This plot shows that plotting the number of components in the preferred age model on a lag plot can be useful in identifying whether a change in lag is due to a change in the resolution of the inversion method.

5 Case Study: Karnali Apatite FT (AFT) data

To make the studies relevant to real data requires the synthetics to be calibrated against the properties of the dataset under consideration. For example, the synthetics should be conditioned on the measured induced track data. In this case study, we demonstrate how to

510 interrogate a detrital apatite fission-track dataset in order to test whether a specific
511 interpretation is robust.
512
513 We consider a detrital AFT dataset from van der Beek et al. (2006). The samples come from
514 a Mio-Pliocene (<15 Ma) series of detrital sediments (Siwalik group) collected in the Karnali
515 River section, western Nepal, and record the denudation history of the central Himalaya
516 during that time interval. Van der Beek et al. (2006) found that multiple age population
517 models were preferred for the younger samples (depositional age <6.5 Ma), for which they
518 reported P_1 ages, whereas only a single age population could be determined for the older
519 samples, hence they reported central ages for these. The relevant Karnali data are replotted
520 in Figure 8. The central ages of all 7 samples are plotted as hollow circles. For the 4 samples
521 with multiple age populations we have plotted the P_1 as filled squares, P_2 as filled circles, P_3
522 as filled triangles; they are coloured by sample number. Following what we have learnt from
523 the synthetic modelling of lag-time plots, we will test the hypothesis that there is insufficient
524 resolution for the population modelling to resolve multiple populations for the older samples.

(a)



(b)

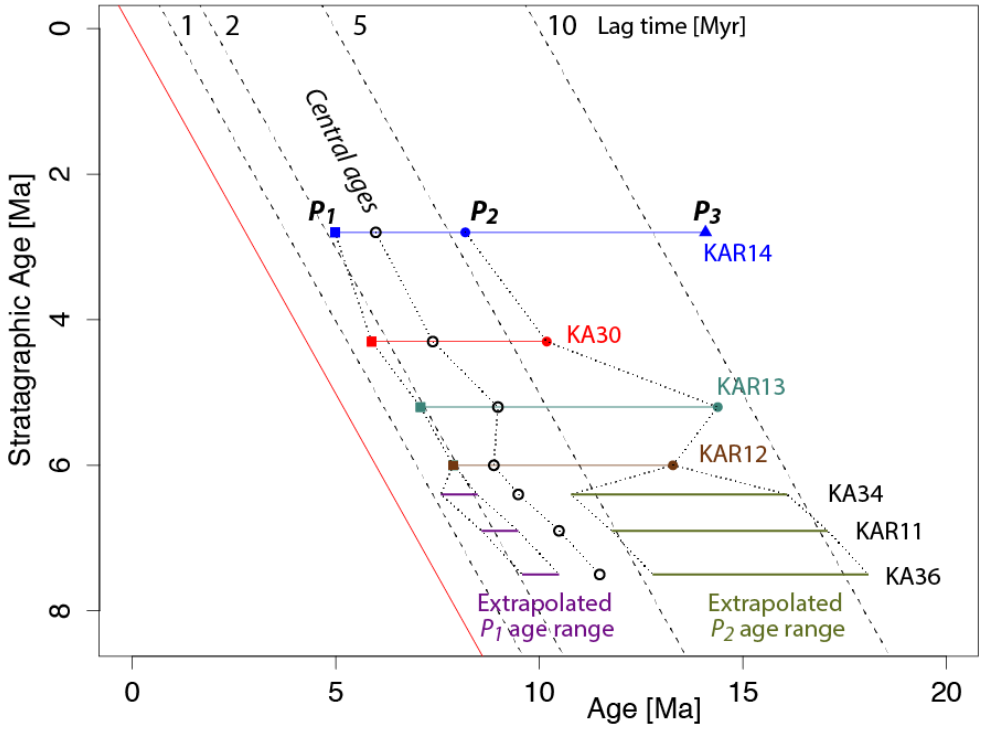


Figure 8 Summary of the un-reset Karnali detrital AFT dataset (Van Der Beek et al., 2006). (a) shows the Karnali data as represented in Figure 8 of the original paper. The younger 4 ages (filled symbols) are P_1 ages and the hollow circles are central ages for the older samples where the parsimonious model only contained a single component. These data appear to show an apparent change in lag at ~6.3Ma, which is coincident with where the number of populations change. (b) All of the population components for the younger 4 samples (filled symbols) and the central ages for all of the samples (hollow circles) are plotted. The extrapolated ranges of where we would expect to see the P_1 and P_2 components to lie have been added for the older samples (horizontal lines, see text for detail).

(a) Young samples with multiple component ages

		Standard Age Analysis [Ma]			Age Calculatations [Ma]	
Sample	nGrains	Stratigraphic	P_1	P_2	$A_{Central}$-P_1	P_2-P_1
KAR14	41	2.8	5	8.2	1	3.2
KA30	41	4.3	5.9	10.2	1.5	4.3
KAR13	16	5.2	7.1	14.4	1.9	7.3
KAR12	44	6	7.9	13.3	1	5.4
				Max	1.9	**7.3**
				Min	1.0	**3.2**

(b) Old samples with a single pooled age

		Standard Age Analysis [Ma]		Age Extrapolations [Ma]	
Sample	nGrains	Stratigraphic	$A_{Central}$	$A_{1,max}$ = $A_{Central}$-1	$A_{1,min}$ = $A_{Central}$-1.9
KA34	63	6.4	9.5	**8.5**	**7.6**
KAR11	51	6.9	10.5	**9.5**	**8.6**
KA36	11	7.5	11.5	**10.5**	**9.6**

Table 1 These tables show the Karnali age data for (a) the young samples with at least 2 ages in the preferred closure age model and (b) the older samples where the preferred solutions contain only a single central age. This data is to be used to test the hypothesis that there is insufficient resolution in the population age modelling method to resolve more than 1 component in the second set of samples purely because they are stratigraphically older. To do this we need to extrapolate the P1 and P2 population ages from the younger samples to the older stratigraphic ages. First, the P1 age in the older samples can be estimated using the range of age differences between the P1 and central ages in the young samples and subtracting these differences from the central ages of the old samples (See bold red values in (b)). Then we can estimate the P2 ages in the older samples by finding the range over which the difference between P1 and P2 varies in the younger samples (See bold red values in (a)). These values have been added to Figure 8.

First, we require a method for finding reasonable estimates of where the A_1 and A_2 ages would lie for the three older Karnali samples. To do this we must make the assumption that the population ages are informative about the true closure ages; we discuss this assumption later. Our estimates are shown as horizontal bars in Figure 8 based on the following reasoning. For the younger samples, we know the central age as well as the P_1 and P_2 population ages. By comparing the P_1 ages with the central ages, we can see that the smallest difference is $1\ \mathrm{Myr}$ and the greatest difference is 1.9 Myr. Our estimates of the A_1 age for the older samples is just their central ages minus 1.0 Myr for the upper limit and 1.9

559 Myr for the lower limit. By comparing the P_1 and P_2 ages for the younger samples, we can
560 find the typical range for ΔP_{12}; the smallest difference is 3.2 Myr and the largest 7.3 Myr. We
561 assume that $\Delta A_{12} \approx \Delta P_{12}$. We choose to use ranges as these will be easy for us to plot in a
562 transparent manner whereas extrapolating a regression of the P_2 ages would be associated
563 with significant uncertainty as the P_2 ages are quite variable.
564
565 Secondly, we need to generate the bootstrapped age models presented in the previous
566 sections (Figures 4-5), but calibrated to each of the 3 older Karnali samples. Since these
567 samples underwent the same analytic process, the main difference between the samples is
568 the number of grains analysed and their uranium content. We perform the bootstrap in the
569 same way as before using a bimodal closure age model. In Figure 9 the preference for 1 or 2
570 peak solutions is plotted as a function of the 2 closure ages (See Supplemental Material for
571 the full age analysis).

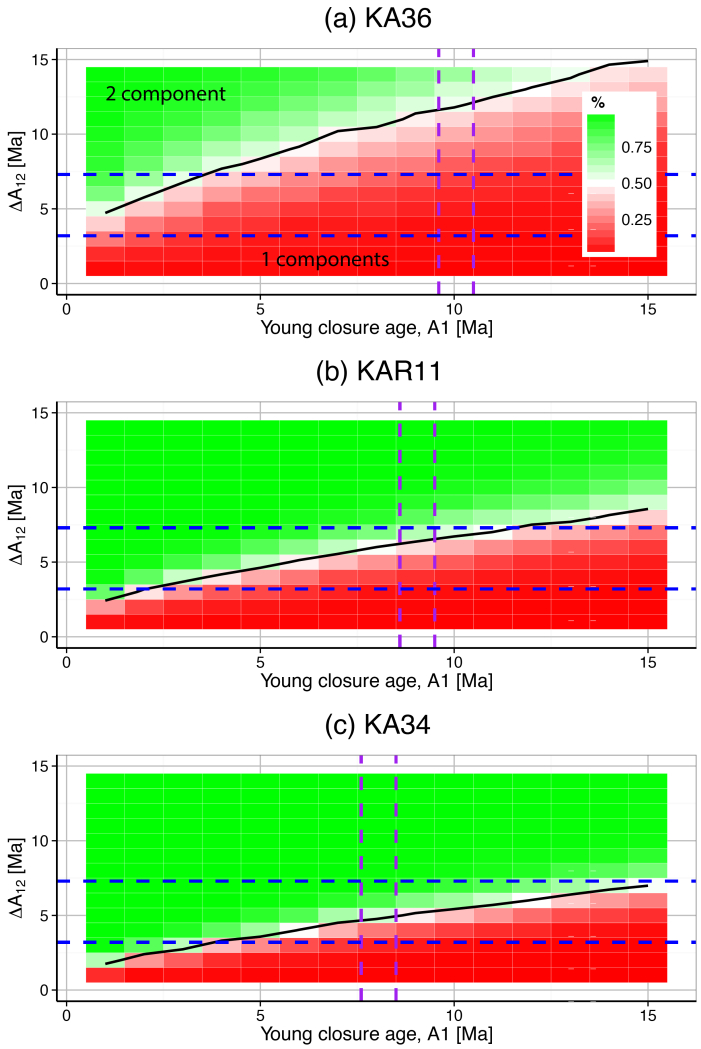


Figure 9 Bootstrapped bimodal closure age models calibrated to the 3 oldest Karnali samples; red indicates a strong preference for single age components in the population modelling and green for 2 components. The vertical bars (purple) show the estimated range for the P1 closure ages and the horizontal bars (blue) show the range for the difference between the youngest and oldest ages based on the data where at least 2 populations are resolved (See Table 1). These figures show that KA36 and KAR11 have a strong preference for single age models and that KA34 has weak preference for bimodal ages but is still consistent with observing a single age model ~40% of the time.

On these figures we have added our estimates for A_1 and ΔA_{12} as vertical and horizontal bars respectively. The KA36 and KAR11 samples have a very strong preference for single population solutions and the preference in KA34 is greater than 40% (Figure 9) despite the true closure age distribution being bimodal. Thus we cannot reject the geologically simpler

584 hypothesis that there has not been a real change in the denudational regime; it can be
585 explained purely in terms of a change in resolution of the data analysis method.
586
587 Note that we cannot say that the modelled populations of the younger samples correspond to
588 the true underlying distributions. Yet, our analysis is not contingent on this as long as we are
589 testing a more specific hypothesis – namely whether the apparent change in lag-time would
590 be resolvable in principle. Whilst we cannot test the properties of the 'true' closure age
591 models as they are unknown in real data, we can probe the ability of the analysis methods to
592 resolve specific features.
593
594 There are other features we could have chosen to analyse, e.g. the youngest sample
595 contains three populations and is the only sample to do so. We could have tested the
596 hypothesis that none of the older samples have the resolution to resolve 3 populations.
597
598 The bootstrap analysis alone does not provide the full solution regarding how to better
599 interpret detrital FT data. What is important is that we set well-defined, testable hypotheses
600 that explore potential limitations of datasets within the geological context in which they are
601 interpreted.

602 # 6 Discussion

603 This study starts to address problems that are known to exist with common practice in the
604 analysis of detrital fission track data. For example, Galbraith (2005) points out that "A finite
605 mixture model is sometimes reasonable in fission-track dating applications when not all
606 grains have the same true age," and goes on to say "However, it will not always be an
607 appropriate model…for heterogeneous ages," and that "the mixing ages may sometimes
608 have a rather artificial meaning." These points are well recognized in the community, yet
609 largely ignored when interpreting data since there have been few frameworks with which to
610 probe the uniqueness of specific interpretations.

611

612 The workflow of many previous studies is to use exhumation scenarios in conjunction with
613 thermal modelling to derive the true closure ages that bedrock and detrital grains would have
614 plausibly experienced (e.g. Bernet and Garver, 2005). These modelled true closure ages are
615 then compared with, for example, the P_1 ages derived from detrital samples. This misses an
616 important step – these forward models did not model observable population ages given some
617 set of true closure ages.

618

619 The bootstrapping of simple, known closure-age models allows us to test hypotheses about
620 whether interpretations are robust. Using a simple bimodal closure-age model, we have
621 shown that statistical artefacts arise in the population modelling that produce apparent
622 changes in the P_1 age when there is no corresponding change in the real closure ages – we
623 can and should routinely diagnose such effects if DFT is to be applied robustly.

624

625 This methodology needs to be considered whatever age inversion algorithms are applied to
626 help interpret detrital data; our intuitive understanding of the data is not yet sufficient to avoid
627 this step. It is beyond the scope of a single paper to present an analysis of all possible
628 models (e.g. minimum age) – but we can summarise our wider findings in that all inversion
629 methods have limitations and assumptions, which produce artefacts that can lead to poor
630 interpretation.

631

632 Further, this analysis allows us to explain why changes in lag-time in detrital AFT data do not
633 necessarily coincide with changes in detrital zircon FT datasets (c.f. Bernet et al., 2006; van
634 der Beek et al., 2006). Since the parameters in the zircon age modelling are different the
635 resolution thresholds will occur at different times and we would expect predictable
636 differences in where lag changes will occur. In particular, the better resolution permitted by
637 detrital zircon FT analysis (due to generally higher N_s and N_i counts, would allow zircon data
638 to better resolve multiple populations for older depositional ages. As an example, detrital

zircon FT data from the same samples as those analysed by van der Beek et al. (2006) resolve 2 age peaks throughout the succession, and resolve 3 age peaks for samples with depositional ages <12 Ma (Bernet et al., 2006).

There is a subtle but important distinction between the concepts of statistical and geological parsimony. Models are often justified using statistical parsimony, which seeks to explain the variance in a dataset with a small number of model parameters; less variance generally only supporting a simpler model. Geological parsimony is more generic but could include concepts such as a geological interpretation requiring a constant lag-time is simpler than one where some external forcing is required to explain a change in lag-time. Tests such as the BIC only attempt to judge using statistical parsimony so caution should be exerted in how to interpret data when a simpler geological model provides a marginally worse statistical explanation of the data. Fundamentally, the tools to trade off statistical and geological parsimony in a routine manner do not exist.

7 Conclusions

Through the application of the Monte-Carlo bootstrap method, we have demonstrated that it is possible to quantify systematic bias and uncertainty in detrital fission track modelling. Understanding these sources of bias is essential if methodologies for rigorously interpreting detrital fission-track data to understand geological process and history are to be developed.

Computationally bootstrapping the behaviour of synthetic DFT samples based on real FT data is an efficient way to improve the interpretation of FT ages and should be a routine part of DFT modelling in order to identify and avoid interpretations that emerge from modelling artefacts. For example, to avoid interpreting changes in P_1 ages as being geologically meaningful in the case where they coincide with statistical artefacts. Through the use of synthetic examples, we have conclusively demonstrated examples of how these artefacts may arise. In the example of the DFT analysis of the Karnali section in Nepal (van der Beek

et al., 2006), we can demonstrate that the apparent decrease in lag-times around 6 Ma can equally be interpreted as a bias in the identification of age populations from older to younger stratigraphic ages.

In comparison to the effort involved in collecting, preparing and analysing DFT samples, this is a fairly painless additional step in the procedure. Given the growing importance being attributed to the interpretation of detrital FT data for inferring climatic and tectonic changes, it is essential that the interpretation of other previous studies are revisited and that the concepts and tools developed in this paper inform future experimental design.

Acknowledgements

MN was funded by a Royal Society of Edinburgh and Scottish Government Personal Research Fellowship. A visit to Grenoble to further collaboration was supported by Marie Curie Actions. For access to the code contact mark.maylor@ed.ac.uk

References

Bernet, M., Brandon, M., Garver, J., Balestieri, M.L., Ventura, B., Zattin, M., 2009. Exhuming the Alps through time: clues from detrital zircon fission-track thermochronology. Basin Research 21, 781–798. doi:10.1111/j.1365-2117.2009.00400.x

Bernet, M., Garver, J.I., 2005. Fission-track Analysis of Detrital Zircon. Reviews in Mineralogy and Geochemistry 58, 205–237. doi:10.2138/rmg.2005.58.8

Bernet, M., van der Beek, P.A., Pik, R., Huyghe, P., Mugnier, J.L., Labrin, E., Szulc, A., 2006. Miocene to Recent exhumation of the central Himalaya determined from combined detrital zircon fission-track and U/Pb analysis of Siwalik sediments, western Nepal. Basin Research 18, 393–412.

Bernet, M., Zattin, M., Garver, J.I., Brandon, M.T., Vance, J.A., 2001. Steady-state exhumation of the European Alps. Geology 29, 35. doi:10.1130/0091-7613(2001)029<0035:SSEOTE>2.0.CO;2

694 Brandon, M., 1992. Decomposition of fission-track grain-age distributions. American Journal
695 of Science.
696 Brandon, M.T., 2002. Decomposition of mixed grain age distributions using Binomfit. On
697 Track 24.
698 England, P., Molnar, P., 1990. Surface uplift, uplift of rocks, and exhumation of rocks.
699 Geology 18, 1173-1177.
700 Fitzgerald, P.G., Sorkhabi, R.B., Redfield, T.F., Stump, E., 1995. Uplift and denudation of the
701 central Alaska Range: A case study in the use of apatite fission track thermochronology
702 to determine absolute uplift parameters. Journal of Geophysical Research 100, 20175–
703 20191. doi:10.1029/95JB02150
704 Fleischer, R.L., Price, P.B., Walker, R.M., 1975. Nuclear Tracks in Solids. Univ of California
705 Press.
706 Galbraith, R.F., 1988. Graphical display of estimates having differing standard errors.
707 Technometrics 30, 271–281.
708 Galbraith, R.F., 2005. Statistics for Fission Track Analysis. Taylor & Francis/CRC Press, 240
709 pp.
710 Galbraith, R.F., Green, P.F., 1990. Estimating the component ages in a finite mixture.
711 Nuclear Tracks and Radiation Measurements 17, 197-206.
712 Galbraith, R.F., Laslett, G.M., 1993. Statistical models for mixed fission track ages. Nuclear
713 Tracks and Radiation Measurements 21, 459–470. doi:10.1016/1359-0189(93)90185-C
714 Gallagher, K., Brown, R., Johnson, C., 1998. Fission track analysis and its applications to
715 geological problems. Annual Review of Earth and Planetary Sciences 26, 519-572.
716 Garver, J.I., Brandon, M.T., 1994. Fission-track ages of detrital zircons from Cretaceous
717 strata, southern British Columbia: Implications for the Baja BC hypothesis. Tectonics 13,
718 401–420. doi:10.1029/93TC02939
719 Garver, J.I., Brandon, M.T., Roden-Tice, M., Kamp, P.J.J., 1999. Exhumation history of
720 orogenic highlands determined by detrital fission-track thermochronology. Geological
721 Society London Special Publications 154, 283–304.

722 doi:10.1144/GSL.SP.1999.154.01.13
723 Gleadow, A.J.W., 1981. Fission-track dating methods: What are the real alternatives?
724 Nuclear Tracks 5, 3–14. doi:10.1016/0191-278X(81)90021-4
725 Glotzbach, C., Bernet, M., van der Beek, P.A., 2011. Detrital thermochronology records
726 changing source areas and steady exhumation in the Western and Central European
727 Alps. Geology 39, 239–242.
728 Hurford, A.J., Green, P., 1983. The zeta age calibration of fission-track dating. Isotope
729 Geoscience 1, 285–317. doi:10.1016/S0009-2541(83)80026-6
730 Jamieson, R.A., Beaumont, C., 1989. Deformation and metamorphism in convergent
731 orogens: a model for uplift and exhumation of metamorphic terrains, in: Daly, J.S., Cliff,
732 R.A., Yardley, B.W.D. (Eds.), Evolution of metamorphic belts. Geological Society Special
733 Publication 43, London, pp. 117-129.
734 Kirstein, L.A., Fellin, M.G., Willett, S.D., Carter, A., Chen, Y.G., Garver, J.I., Lee, D.C., 2010.
735 Pliocene onset of rapid exhumation in Taiwan during arc–continent collision: new
736 insights from detrital thermochronometry. Basin Research 22, 270-285.
737 Price, P.B., Walker, R.M., 1963. Fossil tracks of charged particles in mica and the age of
738 minerals. Journal of Geophysical Research 68, 4847–4862. doi:10.1029/
739 JZ068i016p04847
740 Reiners, P.W., Brandon, M.T., 2006. Using thermochronology to understand orogenic
741 erosion. Annu Rev Earth Planet Sci 34, 419–466, doi: 10.1146/annurev.earth.34.031405.
742 125202
743 Schwarz, G., 1978. Estimating the Dimension of a Model. The Annals of Statistics 6, 461–
744 464. doi:10.1214/aos/1176344136
745 Tagami, T., O'Sullivan, P.B., 2005. Fundamentals of fission-track thermochronology, in:
746 Reiners, P.W., Ehlers, T.A. (Eds.), Low-temperature Thermochronology: Techniques,
747 Interpretation and Application, Reviews in Mineralogy and Geochemistry, 58, pp. 19-47.
748 van der Beek, P., Robert, X., Mugnier, J.-L., Bernet, M., Huyghe, P., Labrin, E., 2006. Late
749 Miocene – Recent exhumation of the central Himalaya and recycling in the foreland

basin assessed by apatite fission-track thermochronology of Siwalik sediments, Nepal. Basin Research 18, 413–434. doi:10.1111/j.1365-2117.2006.00305.x

Vermeesch, P., 2007. Quantitative geomorphology of the White Mountains (California) using detrital apatite fission track thermochronology. J. Geophys. Res. 112, doi: 10.1029/ 2006JF000671.

Vernon, A.J., van der Beek, P.A., Sinclair, H.D., Rahn, M.K., 2008. Increase in late Neogene denudation of the European Alps confirmed by analysis of a fission-track thermochronology database. Earth and Planetary Science Letters 270, 316–329. doi:10.1016/j.epsl.2008.03.053